DOCUMENT RESUME

ED 170 370 TM 008 953

AUTHOR House, Gary D.
TITLE A Comparison of Title I Achievement Results Obtained Under USOE Models A1, C1 and a Mixed Model.
PUB DATE 12 Apr 79
NOTE 32p.; Paper presented at the Annual Meeting of the American Educational Research Association (63rd, San Francisco, California, April 8-12, 1979)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Academic Achievement; *Achievement Gains; Age Differences; Arithmetic; Comparative Statistics; Educationally Disadvantaged; Elementary Education; *Equated Scores; *Evaluation Methods; *Mathematical Models; Program Effectiveness; Reading Comprehension; *Remedial Programs; Research Design; *Test Interpretation
IDENTIFIERS *Elementary Secondary Education Act Title I; Missouri (Saint Louis); Saint Louis Public Schools MO

ABSTRACT

The equivalency of achievement results obtained under Title I evaluation models A1 and C1 was examined. Data were reading comprehension and arithmetic scores on the Iowa Tests of Basic Skills, administered to fourth, sixth, and eighth grade students in St. Louis, Missouri. The St. Louis Public Schools had traditionally used Model A1, a pretest-post test design in which the pretest is not used for selection. Model C1 is the regression-projection model in which only data on students identified as eligible and served by a program are used to assess the impact of Title I programs against a statistical expectation. The expectation is derived from a no-treatment group containing pupils who were eligible, but scored above the cutting score for selection, and were not served by the program. Results indicated, among other findings, that Model C1 produced higher estimates at grade 4, while Model A1 produced higher estimates at grades 6 and 8. The results also raised questions about the selection criteria used to identify program participants. Although the models were not found to be different, definite conclusions could not be drawn about their equivalence. (Two brief post hoc studies, using non-Title I pupils and a mixed model, are appended). (GDC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED170370

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

A Comparison of Title I Achievement Results Obtained Under USOE Models A_1, C_1 and a Mixed Model

By Gary D. House
St. Louis Public Schools

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Gary D. House

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

Paper presented to the Annual Meeting of the American Educational Research Association, at San Francisco, California, on April 12, 1979. Session No. 31.06

IM008 953

PREFACE

This paper was begun as a quick look at the equivalency of achievement results obtained under USOE Proposed Title I Evaluation Models A_1 and C_1. It was originally confined to Title I pupils only. However, once the original study was complete, the author could not resist the temptation to use his data files a bit more. Therefore, the original text is followed by two post hoc studies, briefly described but nevertheless included because they are germane to the question: Should alternate USOE Evaluation Models be implemented interchangeably, and the results obtained from each subsequently aggregated?

The entire paper reads as though it evolved rather than was designed in total and then implemented. It so reads because that is exactly what happened.

Gary D. House

NOTE: Limited numbers of complete appendices containing computer output of data used in this paper are available upon request. They are:

A.	Reading data on Title I eligible pupils:	1. Identified and served; 2. Not identified and not served.
B.	Arithmetic data on Title I eligible pupils:	1. Identified and served; 2. Not identified and not served.
C.	Reading data on pupils not eligible for Title I:	1. Identified; 2. Not identified.
D.	Arithmetic data on pupils not eligible for Title I:	1. Identified; 2. Not identified.

Persons requesting the appendices should specify those which they desire.
Mail inquiries to: Gary House, Divisional Assistant,
Division of Evaluation;
St. Louis Public Schools,
911 Locust Street,
St. Louis, Missouri 63101

INTRODUCTION

General Background

The U.S. Office of Education is expected to issue, momentarily, regulations regarding the reporting of pupil achievement results for ESEA Title I programs.

Some states, including Missouri, have already adopted one or more of the proposed models for use by local school districts. However, other states are still reporting results in any form they choose. There is a great deal of debate occurring among proponents and opponents of the various proposed models. Issues include both technical and practical matters.

Much has been written on the relative legality, precision, comparability, etc. among the models. But what has been written has usually been theoretical or probabilistic. It seems that little real data has been analyzed using different models for the purpose of determining if the models themselves have any bearing on results.

Local Procedures

The St. Louis Public Schools use the adjective "eligible" to indicate that a pupil lives within the geographical boundary of a school which meets the Missouri State economic criterion of deprivation. The adjective "identified" is used to indicate that a pupil meets the educational criterion for deprivation for his grade level.

The approximately 80 schools in St. Louis which meet the Title I economic criterion contain more pupils who meet the educational criterion for their grade level than there are resources to serve them in either the reading or the arithmetic programs which are offered. However, because of

regulations which allow for the extension of remedial services beyond the first year for pupils who have come to achieve beyond the Title I educational criterion, but not yet up to the norm, some pupils who exceed the criterion are legitimately served by the reading program. Other pupils whose achievement levels exceed identification standards are sometimes served because of pupil mobility, teacher judgement or other factors.

Thus, there are eligible pupils scoring at or below identification criteria who are served by Title I remedial reading or arithmetic programs, and there are eligible pupils scoring at or below identification criteria who are not served. Likewise, there are eligible pupils scoring above selection criteria who are not served, and there are eligible pupils scoring above selection criteria who are served.

Reading Comprehension and Arithmetic Subscales of the Iowa Tests of Basic Skills are administered each Fall and Spring to all St. Louis Public School pupils in Grades 3-8. The Spring test serves as a post test for the current academic year, and as a selection test for the following academic year. The Fall test serves only as a pretest for the current year.

Purpose of Study

Although the St. Louis Public Schools have reported Title I evaluation of achievement data using Model A_1 since 1977-78, it is possible that Model C_1 could be used, provided that the two models produce reasonably equivalent and equally valid results.

The purpose of this paper is to describe results obtained by analzying the same pupil achievement data in reading and arithmetic, using proposed models A_1 and C_1 at three different grade levels. The question to be answered is: Do the two models produce different estimates of treatment effect?

For the purposes of this study, Model A_1 is a pretest-post test design in which the pretest is not used for selection. Model C_1 is the regression-projection model in which only data on pupils scoring at or below the program identification cut-off and served by a program are used to assess the impact of the Title I programs against a statistical expectation. That expectation is derived from a no treatment group containing only those Title I eligible pupils who scored above the selection cut-off, and who were not served by the program.[1]

---

[1] Since the audience for this paper is expected to be conversant with all proposed ESEA Title I Evaluation Models, those models are described only superficially. For a more detailed account, see either: Horst, D.P., Tallmadge, G.K., and Wood, C.T. A Practical Guide to Measuring Project Impact on Student Achievement. RMC Research Corporation, Mountain View, California, 1975. Government Printing Office Stock Number 017-080-01460; or, Tallmadge, G.K. and Horst, D.P., A Procedural Guide for Validating Achievement Gains in Educational Projects. RMC Research Corporation, Mountain View, California, 1976.

Limitations

The current study is confined to data from Grades 4, 6 and 8 because of time limitations, because these grades represent a range of middle and upper level enrollments, and because all data used are from a single test battery. Although these conditions produce the benefits of time conservation, representativeness of populations and a high probability of inter-test correlation, they also result in a few handicaps.

Reading and arithmetic programs in St. Louis have heavy primary level enrollments. At the primary levels (Grades 1-3), several different tests are used for pupil selection, and no pretests are administered. Although unavoidable under these conditions, it must be stated that this study does not represent the entire range of possible program results.

Another limitation of confining this study to three grades is a loss of statistical power in comparing model results. It would have been preferable to look at eight grades instead of only three; however, that luxury was not available.

METHODS

1. A file of Iowa Test of Basic Skills Reading Comprehension and Total Arithmetic Subtest normal curve equivalent scores from Spring, 1977. Fall, 1977 and Spring, 1978 was created, including all St. Louis Public School Pupils who were in Grades 4, 6, or 8 during 1977-78. Also included in the file were flags as to each pupil's participation or non-participation in both a Title I Remedial Reading and/or Remedial Arithmetic program.
2. Model A_1 evaluation data for each grade and program were derived by first selecting those pupils, served by the programs, who were identified for such services according to their Spring, 1977 test scores (N.C.E. $\leq$ 45 for Grades 4 and 6, and $\leq$ 42 for Grade 8). The mean Fall, 1977 (pretest) and Spring, 1978 (post test) N.C.E. scores on the appropriate subtests were then computed, eliminating all missing values. The mean N.C.E. gain for each group was obtained by subtracting the mean Fall, 1977 scores from the mean Spring, 1978 scores.
3. Model C_1 evaluation data for each grade and program were derived by first selecting those eligible pupils who scored above their appropriate identification criterion on the corresponding Spring, 1977 ITBS Subtest and who were not served by Title I during 1977-78. Scatterplots of Spring, 1977 and Spring, 1978 subtest scores were made, while calculating the corresponding R, R^2, B (slope) and I (intercept) for each set.

The expected post test means for the treatment groups were calculated using the formula: $\hat{\bar{Y}}_t = \bar{Y}_c + b_c (\bar{X}_t - \bar{X}_c)$

where $\bar{Y}_c$ = actual post test mean of comparison group

b_c = slope of selection test versus post test-comparison group

$\bar{X}_t$ = actual pretest mean for treatment group

$\bar{X}_c$ = actual pretest mean for comparison group

Model C_1 treatment effect was calculated by subtracting the expected post test treatment group mean from the actual treatment group post test mean: Effect $_t = \bar{Y}_t - \hat{\bar{Y}}_t$

4. Selection test, pretest and post test means, and differences between these sequential means for each group, were calculated.

5. To judge whether or not Model A_1 versus Model C_1 results were different across all six groups (3 reading and 3 arithmetic grades), the binomial sign test was employed, with an alpha level of .10 selected for the rejection of the null hypothesis.

FINDINGS

Means for each test administration and differences between successive means for each group are shown in Table 1.

TABLE 1

MEAN AND DIFFERENCE N.C.E. READING AND MATH SCORES AMONG SELECTION, PRETEST AND POST-TEST PERIODS FOR TREATMENT AND COMPARISON GROUPS

	Selection	D_1	Pretest	D_2	Post-Test	N
			Reading			
			Grade 4			
T	26.72	-.42	26.30	-.22	26.08	1408
C	56.29	-13.99	42.30	+ 2.74	45.04	934
			Grade 6			
T	27.96	+ .49	28.45	+ 2.32	30.77	1806
C	55.30	- 9.71	45.59	+ 3.15	48.74	559
			Grade 8			
T	27.62	-4.64	22.98	+ 4.18	27.16	1159
C	53.13	-10.75	42.38	+ 5.19	47.57	877
			Math			
			Grade 4			
T	24.38	+ 3.26	27.64	-.61	27.03	685
C	59.19	-14.67	44.52	+ 2.30	46.82	1106
			Grade 6			
T	26.70	+ 1.01	27.71	+ 3.52	31.23	789
C	57.87	-13.73	44.14	+ 5.44	49.58	1032
			Grade 8			
T	26.63	-1.37	25.26	+ 6.65	31.91	485
C	54.51	-8.89	45.62	+ 5.51	51.13	1044

Table 1 shows substantial reductions in means from selection to pretest periods for comparison group pupils, across grade levels, and across subtests. Although post test results for these pupils were somewhat higher than pretest results, they did not regain their selection test levels in any instance.

Treatment group pretest means were slightly higher than their selection test means at Grade 6 in reading, and Grades 4 and 6 in math. Slight reductions in their pretest means occurred elsewhere.

Pretest to post test means showed positive change in both reading and math for all groups except Grade 4 treatment.

All math treatment groups demonstrated some positive movement against their norming populations across selection, pretest and post test periods. Comparison groups had net losses in math during the same time frame. In reading, all except the Grade 6 treatment group acquired net losses. (The Grade 6 reading treatment group showed net losses. The Grade 6 reading treatment group obtained a net gain of 2.81 N.C.E. or, .133 N.C.E. standard deviations).

Table 2 presents Model A_1 versus Model C_1 results for Reading Treatment Groups. Also shown are selection test (Spring, 1977) to post test (Spring, 1978) regression slopes for treatment and comparison pupils.

TABLE 2

MODEL A_1 VERSUS MODEL C_1
RESULTS FOR READING TREATMENT GROUPS
AT THREE GRADE LEVELS

MODEL A_1 RESULTS		MODEL C_1 RESULTS	
GRADE 4			
Pretest Mean	= 26.3	Predicted Post Mean	= 22.6
Post-test Mean	= 26.1	Observed Post Mean	= 26.1
Post - Pre Mean	= - 0.2	Observed - Predicted Mean	= 3.5
GRADE 6			
Pretest Mean	= 28.4	Predicted Post Mean	= 29.3
Post-test Mean	= 30.7	Observed Post Mean	= 30.7
Post - Pre Mean	= 2.3	Observed - Predicted Mean	= 1.4
GRADE 8			
Pretest Mean	= 23.0	Predicted Post Mean	= 24.6
Post-test Mean	= 27.1	Observed Post Mean	= 27.1
Post - Pre Mean	= 4.1	Observed - Predicted Mean	= 2.5

Selection Test Versus Post-test Slopes for Title I
Reading Treatment Versus Comparison Groups

	Treatment	Comparison
Grade 4	.29	.76
6	.46	.71
8	.45	.90

Model C_1 produced a slightly higher estimate of reading treatment effect at Grade 4 than did Model A_1. At Grades 6 and 8, the reverse was true. The absolute magnitude of the differences in treatment effect estimation was 3.3 N.C.E. at Grade 4, 1.1 N.C.E. at Grade 6, and 1.6 N.C.E. at Grade 8.

Treatment versus Comparison group Selection test to post test regression slopes were substantially different across the grades.

Model Comparisons for the Math Program are presented in Table 3.

TABLE 3

MODEL A_1 VERSUS MODEL C_1
RESULTS FOR MATH TREATMENT GROUPS
AT THREE GRADE LEVELS

MODEL A_1 RESULTS		MODEL C_1 RESULTS	
GRADE 4			
Pretest Mean	= 27.6	Predicted Post Mean	= 20.7
Post-test Mean	= 27.0	Observed Post Mean	= 27.0
Post - Pre Mean	= - 0.6	Observed - Predicted Mean	= 6.3
GRADE 6			
Pretest Mean	= 27.7	Predicted Post Mean	= 27.8
Post-test Mean	= 31.2	Observed Post Mean	= 31.2
Post - Pre Mean	= 3.5	Observed - Predicted Mean	= 3.4
GRADE 8			
Pretest Mean	= 25.3	Predicted Post Mean	= 26.3
Post-test Mean	= 31.9	Observed Post Mean	= 31.9
Post - Pre Mean	= 6.6	Observed - Predicted Mean	= 5.6

Comparison of Model C_1 Slopes for Title I Math Treatment
Versus Comparison Groups

	Treatment	Comparison
Grade 4	.39	.75
6	.46	.70
8	.50	.89

Table 3 indicates that Model C_1 produced lower estimates of treatment effect than did Model A_1 at Grades 4 and 6. The absolute magnitude of the differences between model results was .1 N.C.E. and 1.0 N.C.E. respectively at these grades. Application of Model C_1 to Grade 4 math data resulted in an estimate of treatment effect 6.9 N.C.E. higher than that of Model A_1.

As was true of the reading data, regression slopes for treatment and comparison groups were different.

Table 4 presents the results of a test of differences between treatment effect estimates obtained under the two ESEA proposed models.

TABLE 4

Binomial Sign Test for Comparison of Model A_1 and Model C_1 Estimates of Title I Treatment Effect (+ = higher estimates of effect)

	Model A_1	Model C_1
Grade 4 Reading		+
Grade 6 Reading	+	
Grade 8 Reading	+	
Grade 4 Math		+
Grade 6 Math	+	
Grade 8 Math	+	

n=6; r=4; p=.688

The binomial sign test indicated that the two-tailed probability of one model producing different estimates of treatment effects four out of six possible times by chance alone was 68.8%. Therefore, the hypothesis of no difference was accepted.

DISCUSSION AND RECOMMENDATIONS

This study is limited to two programs at three grade levels in one large urban area. As such, it probably provides more questions than answers; yet, it contributes some interesting and useful information:

1. Although not consistantly different estimates of treatment effect were obtained under the two models, it is puzzling as to why Model C_1 produced higher estimates for both subject matter areas at Grade 4, while Model A_1 produced higher estimates at Grades 6 and 8. An expanded study throughout the grades could yield information which would be more conclusive. Puzzling also is that only at Grade 4 were pretest to post test results in both reading and arithmetic indicative of (very slight) losses in relative position within the national norms for treatment groups.
2. Shifts in the distributions of selection test versus pretest administrations were quite different for treated pupils versus those whose selection scores were too high for treatment. Treatment group selection to pretest mean differences were as many times positive as negative (contrary to regression theory), and were minor in magnitude. On the other hand, each of the comparison groups had dramatic negative shifts between these two testing periods: In reading, these reductions were .66, .46 and .51 N.C.E. standard deviations at Grades 4, 6 and 8 respectively. In arithmetic, they were .69, .65 and .42 N.C.E. standard deviation units lower.

These data cannot be explained by regression effects alone, and certainly indicate a need to re-examine the selection process for Title I eligible pupils. It seems that most pupils originally selected would continue to be identified for services, given Fall test results. However, many pupils scoring too high for program selection in the Spring would be well within selection criteria if Fall tests were used for selection purposes.

3. The large differences in selection to post test regression slopes brings into serious question the validity of the Model C_1 assumption of homogeneity of regression for comparison versus treatment populations. It is recommended that an identical analysis of similar data on Non Title I eligible pupils be conducted. Such a study could provide a test for equivalent slopes between "At or Below" identification criteria pupils versus pupils scoring "Above" identification criteria.

   It would also be interesting to see whether the two models (A_1 versus C_1) would produce equivalent estimates of learning for the "At or Below" group, given no treatment. If the two models are in fact equally valid methods of estimating learning, then, for a no treatment population, it would seem logical to expect equivalent results.

   It is recommended that the St. Louis Public Schools undertake a careful study of their testing procedures in regard to both the identification of pupils for Title I treatment and the subsequent evaluation of that treatment.

   Further, it seems premature to draw any conclusions regarding the equivalence of Models A_1 and C_1. This study did not find them different. Yet it indicated that its failure to find differences may have been due to its limitations.

Before any state or national group having decision-making power over local education agencies mandates that the evaluation of Title I programs can proceed in multiple formats, but can later be aggregated and/or compared, it is strongly urged that much more empirical evidence on model equivalency be obtained.

TWO POST HOC STUDIES

1) A Reproduction of Preceding Study Using Data on Non Title I Pupils

2) The Exploratory Application of a Mixed Model

POST HOC STUDY NO. 1

A Reproduction of the Preceding Study using Data on Non Title I Pupils

Explanation:

The preceding study was duplicated exactly, except for use of data from pupils who were not eligible for Title I services. In this study comparing the estimates of learning obtained under Model A_1 versus Model C_1 procedures, the treatment group is replaced with those ineligible pupils scoring "At or Below" criteria levels,and the comparison group is replaced with ineligible pupils scoring "Above" criteria levels.

Table presentations follow in the same sequence and formats as those in the first study. Each table is numbered as were those in the original study, but with subscripts.

TABLE 1-A

MEAN AND DIFFERENCE N.C.E. READING AND MATH SCORES AMONG SELECTION, PRETEST AND POST-TEST PERIODS FOR NON-TITLE I STUDENTS SCORING AT OR BELOW VERSUS ABOVE THE SELECTION CRITERIA

Reading

Grade 4

	Selection	D_1	Pretest	D_2	Post-Test	N
At or Below	28.58	+ 2.67	31.25	- 0.57	30.68	547
Above	60.92	- 8.49	52.43	+ 0.33	52.76	809

Grade 6

At or Below	30.12	+ 2.25	32.37	+ 0.68	33.05	761
Above	57.65	- 0.67	56.98	+ 0.86	57.84	774

Grade 8

At or Below	29.57	+ 1.05	30.62	+ 1.71	32.33	725
Above	58.53	- 2.53	56.00	+ 2.48	58.48	1059

Math

Grade 4

At or Below	29.98	+ 4.38	32.36	- 0.92	31.44	505
Above	63.59	- 9.45	54.14	+ 0.11	54.25	837

Grade 6

At or Below	29.80	+ 2.65	32.45	+ 0.71	33.16	635
Above	64.31	- 7.89	56.42	+ 1.69	58.11	897

Grade 8

At or Below	29.89	+ 2.45	32.34	+ 1.26	33.60	7[illegible]
Above	60.07	- 2.35	57.72	+ 2.82	60.54	1064

TABLE 2-A
MODEL A_1 VERSUS MODEL C_1 READING STATISTICS
FOR NON-TITLE I STUDENTS SCORING AT OR BELOW SELECTION
CRITERIA AT THREE GRADE LEVELS

MODEL A_1 RESULTS			MODEL C_1 RESULTS		
			GRADE 4		
Pretest Mean	=	31.2	Predicted Post Mean	=	22.7
Post-test Mean	=	30.7	Observed Post Mean	=	30.7
Post - Pre Mean	=	- 0.5	Observed - Predicted Mean	=	8.0
			GRADE 6		
Pretest Mean	=	32.3	Predicted Post Mean	=	32.2
Post-test Mean	=	33.0	Observed Post Mean	=	33.0
Post - Pre Mean	=	0.7	Observed - Predicted Mean	=	0.8
			GRADE 8		
Pretest Mean	=	30.6	Predicted Post Mean	=	29.8
Post-test Mean	=	32.3	Observed Post Mean	=	32.3
Post - Pre Mean	=	1.7	Observed - Predicted Mean	=	2.5

Selection Test Versus Post-test Slopes for Non Title I
At or Below Criterion Versus Above Criterion Groups (Reading)

	At or Below	Above
Grade 4	.34	.93
6	.51	.93
8	.42	1.00

TABLE 3-A
MODEL A_1 VERSUS MODEL C_1 MATH STATISTICS
FOR NON-TITLE I STUDENTS SCORING AT OR
BELOW SELECTION CRITERIA FOR THREE GRADE LEVELS

MODEL A_1 RESULTS

GRADE 4

Pretest Mean	=	32.4	Predicted Post Mean =	25.7
Post-test Mean	=	31.4	Observed Post Mean =	31.4
Post - Pre Mean	=	-1.0	Observed - Predicted Mean =	5.7

GRADE 6

Pretest Mean	=	32.4	Predicted Post Mean =	24.3
Post-test Mean	=	33.2	Observed Post Mean =	33.2
Post - Pre Mean	=	- 0.8	Observed - Predicted Mean =	8.9

GRADE 8

Pretest Mean	=	32.3	Predicted Post Mean =	29.8
Post-test Mean	=	33.6	Observed Post Mean =	33.6
Post - Pre Mean	=	1.3	Observed - Predicted Mean =	3.8

Selection Test Versus Post-test Slopes for Non-Title I
At or Below Criterion Versus Above (Math) Criterion Groups

	At or Below	Above
Grade 4	.39	.85
6	.47	.98
8	.56	1.02

TABLE 4-A

Binomial Sign Test for Comparison of Model A_1 and Model C_1 Estimates of Learning for Non Title I Eligible Pupils
(+= higher estimate of learning)

	Model A_1	Model C_1
Grade 4 Reading		+
Grade 6 Reading		+
Grade 8 Reading		+
Grade 4 Math		+
Grade 6 Math		+
Grade 8 Math		+

n=6; r=6; p=.032

DISCUSSION

Some descriptive comparisons between this second set of analyses and the original analyses can be stated. First, all selection to pretest mean differences for "At or Below" criteria versus "Above" criteria groups, are in the direction to be expected from regression theory. The magnitudes of these shifts vary a great deal without explanation.

As was true in the original study, only Grade 4 pretest to post test mean differences were negative. This may suggest a problem within the Grade 4 testing program.

Unlike the original study, this one showed Model C_1 as producing higher estimates of learning than did Model A_1, across all groups.

Interestingly, regression slopes for the "Above" criteria groups were higher than for the parallel Title I comparison groups, but the slopes for the "At or Below" groups were very similar to those of the Title I treatment groups. This finding raises the possibility of comparing Title I treatment group achievement results to those of Non Title I pupils scoring similarly on the selection measure.

Although no firm conclusions regarding the relative equivalence of Models A_1 and C_1 are yet justified, the data from this second study strongly suggest that results may be dependent upon populations as well as models. Clearly, the problems of model selection on the part of local education agencies, and of model mandates on the part of others, are not solved.

POST HOC STUDY NO. 2

An Exploratory Application of a Mixed Model

Explanation:

Because Post Hoc Study No. 1 showed roughly equivalent selection to post test slopes for "At or Below" Non Title I pupils, as compared to those found in the original study for Title I treatment pupils, it was decided that a mixed evaluation model might be justified.

In this model, the post test achievement means (adjusted for pretest differences through analysis of covariance procedures) of Title I treatment versus Non Title I "At or Below" pupils were compared.

An alpha level of .05 was selected for a two-tailed F test of mean differences.

Results are summarized in the following sets of tables:

ANALYSIS OF COVARIANCE
AND DESCRIPTIVE STATISTICS TABLES
FOR READING TREATMENT AND COMPARISON GROUPS

GRADE 4

Source of Variation	SS	DF	MS	F	Alpha
Pretest (Covariate)	35210.18	1	35210.18	225.96	.001
Reading Treatment	3667.81	1	3667.81	23.54	.001
Explained	43216.75	2	21608.38	138.68	.001
Error	288878.94	1854	155.81		
Total	332095.69	1856	178.93		

Group	N	Unadjusted Post Test Means	Adjusted Post Test Means
Treatment	1372	26.36	26.51
Comparison	485	30.85	29.39

GRADE 6

Source of Variation	SS	DF	MS	F	Alpha
Pretest (Covariate)	78956.38	1	78956.38	550.52	.001
Reading Treatment	448.20	1	448.20	3.12	N.S.
Explained	82420.44	2	41210.22	287.34	.001
Error	329153.00	2295	143.42		
Total	411573.44	2297	179.18		

Group	N	Unadjusted Post Test Means	Adjusted Post Test Means
Treatment	1644	30.73	31.38
Comparison	654	33.45	31.21

GRADE 8

Source of Variation	SS	DF	MS	F	Alpha
Pretest (Covariate)	72148.56	1	72148.56	560.10	.001
Reading Treatment	2421.45	1	2421.45	18.80	.001
Explained	87702.00	2	43851.00	340.42	.001
Error	228001.12	1770	128.81		
Total	315703.12	1772	178.16		

Group	N	Unadjusted Post Test Means	Adjusted Post Test Means
Treatment	1152	27.15	28.43
Comparison	621	33.35	30.97

ANALYSIS OF COVARIANCE
AND DESCRIPTIVE STATISTICS TABLE
FOR MATH TREATMENT AND COMPARISON GROUPS

GRADE 4

Source of Variation	SS	DF	MS	F	Alpha
Pretest (Covariance)	32591.55	1	32591.55	181.31	.001
Math Treatment	2606.95	1	2606.95	14.50	.001
Explained	39242.44	2	19621.22	109.15	.001
Error	198452.44	1104	179.76		
Total	237694.88	1106	214.91		

Group	N	Unadjusted Post Test Means	Adjusted Post Test Means
Treatment	660	27.02	27.76
Comparison	447	32.02	30.93

GRADE 6

Source of Variation	SS	DF	MS	F	Alpha
Pretest (Covariate)	26375.10	1	26375.10	150.43	.001
Math Treatment	88.58	1	88.58	0.51	N.S.
Explained	27867.06	2	13933.53	79.47	.001
Error	230040.62	1312	175.34		
Total	257907.69	1314	196.28		

Group	N	Unadjusted Post Test Means	Adjusted Post Test Means
Treatment	766	31.30	31.98
Comparison	549	33.46	32.51

GRADE 8

Source of Variation	SS	DF	MS	F	Alpha
Pretest (Covariate)	49540.43	1	49540.43	334.26	.001
Math Treatment	75.08	1	75.08	0.51	N.S.
Explained	51806.38	2	25903.19	174.77	.001
Error	160216.62	1081	148.21		
Total	212023.00	1083	195.77		

Group	N	Unadjusted Post Test Means	Adjusted Post Test Means
Treatment	478	31.89	33.31
Comparison	606	34.80	33.28

RESULTS

In this mixed model, no matching and no tests of differences between pretest levels were attempted.

Significant differences were found in three of six statistical tests. In each instance of difference, Title I groups came out on the lower end of the achievement continuum.

DISCUSSION

Looking across all three of the studies contained in this paper, one must be impressed with a lack of firm evidence as to which direction should be taken in evaluating achievement effects for Title I programs. Each attempt to look at alternative models has been done with the knowledge that no matter what the procedure, the inherent problems associated with it almost preclude its use. Seemingly, the purpose and goals of Title I make severe restrictions on the design of evaluation models while at the same time providing for an almost infinite number of ways in which those restrictions can be met. Only a few have been studied here.

The evidence is strong that no uniform evaluation process will ever be possible for implementation nationwide, such that results from all local education agencies are aggregatable. To the author, it seems ludicrous to continue to try. It seems much more sensible to stop trying the impossible than to continue the nationwide debate forever.

It would be far less expensive and far more illuminating for the U.S. Office of Education to contract for a controlled study of Title I effects nationwide by a neutral, but competent, research unit which could have the

mandated cooperation of selected local educational agencies, than it will ever be to accomplish what they are currently trying to do.

If they are not willing to obtain such control, then they should probably acknowledge that good information about Title I effects will never be available.

APPENDICES

A.	Reading data on Title I eligible pupils:	Identified and served Not identified and not served
B.	Arithmetic data on Title I eligible pupils:	Identified and served Not identified and not served
C.	Reading data on pupils not eligible for Title I:	Identified Not identified
D.	Arithmetic data on pupils not eligible for Title I:	Identified Not identified